



Proyecto Análisis de Datos Explorador

Exploración de datos del programa DigiCampus: Acceso, cobertura y beneficios

**Juan José Restrepo Tobón
Neila Sirley Perilla Perdomo
Yonatan Alberto López Sierra
Andrés Felipe Jaramillo Marén**

**2025
Julio**

1. Introducción	3
2. Identificación del problema	3
3. Metodología	5
4. Pasos implementados desagregadamente.	6
Paso 1	8
Paso 2	8
Paso 3	9
Paso 4	10
Paso 5	10
Paso 6	11
Paso 7	12
5. Hallazgos	13
Distribución de los beneficios	14
Clasificación académica	15
Instituciones a las que ingresan	16
Modalidad de estudio	17
Municipios de donde provienen la mayoría de estudiantes	19
Alcance de la política	20
Relación incentivo, Género y Ciudad.	21
Conclusiones	21
Reflexiones	22

1. Introducción

El acceso a la educación superior continúa siendo uno de los principales retos de las políticas públicas en Colombia, especialmente en contextos marcados por desigualdades económicas, geográficas y de género. En respuesta a este desafío, la Gobernación del Valle del Cauca implementó el programa *DigiCampus*, una iniciativa orientada a ampliar las oportunidades educativas mediante el otorgamiento de descuentos parciales y totales en matrículas de instituciones de educación superior (IES) de la región. Esta política busca reducir las barreras de ingreso a la universidad, mejorar los índices de cobertura y fomentar la permanencia estudiantil, especialmente entre poblaciones tradicionalmente excluidas.

No obstante, aunque la cobertura del programa ha sido significativa, aún persisten preguntas fundamentales sobre los criterios y patrones de asignación de los beneficios. En particular, es necesario indagar si existen diferencias significativas entre hombres y mujeres en el acceso a becas completas y parciales, lo que permitiría identificar posibles sesgos o desigualdades de género en la distribución de los recursos públicos.

Este proyecto de investigación tiene como propósito cuantificar y comparar el acceso a los beneficios de *DigiCampus* entre mujeres y hombres, analizando también cómo estas diferencias se expresan según otras variables como la universidad, el municipio de residencia y la modalidad de estudio. A través del análisis estadístico de la base de datos de beneficiarios, se espera aportar evidencia empírica que contribuya al seguimiento, evaluación y mejora del programa desde un enfoque de equidad.

2. Identificación del problema

El objetivo central de *DigiCampus* es ampliar las oportunidades de acceso a la educación superior con pertinencia territorial y equidad social. No se trata únicamente de aumentar la cobertura, sino de garantizar una oferta formativa articulada con las necesidades y vocaciones productivas de cada municipio, integrando además componentes de innovación pedagógica, infraestructura tecnológica y acompañamiento comunitario. Así, se busca generar arraigo en los jóvenes, fortalecer los tejidos sociales locales y dinamizar el desarrollo económico regional desde la educación.

Entre sus metas específicas, se destacan:

- a) Reducir las barreras de ingreso a la educación superior, especialmente para mujeres, comunidades afrodescendientes, víctimas del conflicto armado y población rural.
- b) Promover un modelo de formación híbrido y multimodal, que combine presencialidad, virtualidad y prácticas situadas, adaptándose a las realidades del territorio.
- c) Dotar a los municipios de infraestructura digital, como espacios públicos conectados, que permitan a los estudiantes acceder a recursos tecnológicos aunque no los tengan en casa.
- d) Fomentar competencias digitales críticas y creativas, más allá del simple acceso a internet, desarrollando capacidades para el pensamiento computacional, el análisis de datos y el uso ético de la tecnología.

Se espera que, a través de sus distintas convocatorias, la política pública DigiCampus logre beneficiar a miles de estudiantes, garantizando una distribución equitativa entre becas completas y parciales, y priorizando la inclusión de poblaciones históricamente excluidas. A mediano y largo plazo, se proyecta que esta iniciativa contribuya a reducir las tasas de deserción educativa, fortalecer las trayectorias formativas locales y cerrar las brechas territoriales en el acceso a la educación superior, consolidándose como un modelo replicable, sostenible y con impacto estructural en el desarrollo regional.

La problemática a abordar en la presente investigación es la revisión de los impactos de ese programa. Identificar si ha logrado los objetivos, en qué medidas los ha logrado y si en efecto hay una democratización por medio de tal política pública.

Pregunta problema

¿Existen diferencias significativas entre hombres y mujeres en el acceso a becas completas y parciales dentro del programa DigiCampus en el Valle del Cauca?

Objetivo general

En este contexto, el objeto general de esta investigación es analizar el impacto de la política pública DigiCampus en el acceso a la educación superior en el Valle del Cauca, con énfasis en la distribución de beneficios según género, para identificar posibles brechas o avances en equidad educativa.

Objetivos específicos

Para lograr ese objetivo general, se disponen los siguientes objetivos específicos

1. Cuantificar cuántas mujeres y hombres han recibido becas completas y parciales.
2. Comparar la proporción de becas completas entre mujeres y hombres.
3. Identificar patrones de asignación de beneficios según sexo y otras variables como universidad, municipio o modalidad.

3. Metodología

La presente investigación adopta un enfoque cuantitativo y exploratorio, centrado en el análisis de datos abiertos para evaluar el impacto de la política pública DigiCampus, particularmente en lo que respecta a la distribución de becas según género. El eje metodológico principal es la recolección, limpieza y visualización de datos, con el objetivo de identificar patrones, tendencias y posibles desigualdades en el acceso a los beneficios educativos ofrecidos por el programa.

La primera fase consistió en la búsqueda y recopilación de información en fuentes oficiales, principalmente en los portales institucionales de la Gobernación del Valle del Cauca y entidades relacionadas con la implementación de DigiCampus. Los datos seleccionados fueron evaluados en términos de pertinencia, actualidad y grado de desagregación, dando prioridad a aquellos que permitieran distinguir entre tipos de beca (completa o parcial), características sociodemográficas y, especialmente, la variable de género.

Una vez identificadas las bases de datos más relevantes, estas fueron importadas al entorno de programación Python. Desde allí se desarrolló un proceso de limpieza y depuración, que incluyó la eliminación de registros incompletos, estandarización de variables y filtrado de

observaciones no pertinentes para los objetivos del estudio. Posteriormente, se procedió a la organización de las variables clave en función del tipo de análisis requerido.

Para la etapa de análisis, se aplicaron herramientas de visualización de datos y estadística descriptiva con apoyo de bibliotecas como *pandas*, *matplotlib* y *seaborn*. Esto permitió generar gráficos comparativos, distribuciones de frecuencia y relaciones proporcionales que ilustran las diferencias (o similitudes) en la asignación de beneficios entre mujeres y hombres. Estos procedimientos fueron realizados con base en metodologías aprendidas previamente en un curso de análisis de datos, garantizando la validez técnica de las representaciones gráficas y la interpretación de resultados.

El análisis no se limita a la simple cuantificación, sino que busca identificar posibles patrones de inequidad de género en la distribución de beneficios, así como variaciones asociadas a otras variables como el municipio, la modalidad educativa o el tipo de institución. En este sentido, la metodología propuesta permite aproximarse con solidez al objetivo general de la investigación y ofrece bases empíricas para futuras evaluaciones de impacto.

4. Pasos implementados desagregadamente.

En un primer momento se importan las librerías óptimas para un análisis de datos:

```
• import pandas as pd
• import matplotlib.pyplot as plt
• import seaborn as sns
•
```

Luego de descargada la base de datos procedemos a vincularla con google colab, para poder trabajar con ella. Así mismo imprimimos las primeras líneas del archivo para tener nociones de la organización de la base de datos con *print(data.info())*

```
# Lectura

from google.colab import drive

#drive.mount('/gdrive')
```

```

ruta =
"https://raw.githubusercontent.com/ylopezs/proyecto_analisis/refs/heads/main/Apropiaci_n_
TIC_del_Valle_del_Cauca_20250625.csv"

data = pd.read_csv(ruta)

print(data.head())

print(data.info())

```

Seguidamente queremos limpiar la base de datos y procedemos a indicarle al programa que nos señale las casillas vacías y el tipo de objeto de las columnas

```

data.dropna(inplace=True)

print(data.info())

```

Ahora, de acuerdo con nuestros objetivos y necesidades, seleccionamos las columnas que consideramos relevantes

```

columnas_relevantes = ['Municipio','Sexo','Clasificacion Academica','Entidad
Organización','Programa','Modalidad','Beneficio', 'Latitud', 'Longitud']

for col in columnas_relevantes:

print(f'Columna {col}: {data[col].nunique()} subniveles')

```

Luego, procedemos a solicitar una visualización de todos aquellos datos de utilidad por medio de gráficos de barras

```

fig, ax= plt.subplots(nrows=7, ncols=1, figsize=(10,30))

fig.subplots_adjust(hspace=1)

```

```
for i, col in enumerate(columnas_relevantes):  
  
    sns.countplot(x=col, data=data, ax=ax[i])  
  
    ax[i].set_title(col)  
  
    ax[i].set_xticklabels(ax[i].get_xticklabels(), rotation = 90)
```

Posteriormente procedemos a desagregar los datos para obtener todo aquello que consideramos valioso.

Paso 1

Primero se procede a introducir un código que nos indique tanto el número de mujeres como de hombres, así como otros géneros de ser el caso.

```
sex_counts = data['Sexo'].value_counts()  
  
print(sex_counts)
```

El resultado arrojado fue que hay más mujeres que hombres y que el género no binario es mínimo.

Mujeres: 2238

Hombres: 1828

No binario: 13

Paso 2

Para mayor comodidad visual, deseamos que estos datos se expresen en una gráfica de barras. Para tal propósito introducimos el código

```
sns.countplot(x='Sexo', data=data)
```



```
plt.show()
```

El resultado arrojado por el programa lo adjuntamos en la sección de hallazgos. Por ahora nos limitaremos a exponer los procedimientos de código.

Paso 3

Luego de identificar la cantidad de mujeres y hombres introducimos un gráfico, también de barras, que nos indique la cantidad de mujeres y hombres que son parte del programa, procedemos a identificar qué tipo de beneficio reciben, es decir, procedemos a visualizar si reciben una beca parcial o total, con los siguientes códigos:

Para mujeres:

```
f_beneficio= sns.countplot(data=data[data['Sexo']=='F']['Beneficio'])

plt.xlabel("Cantidad de mujeres ")

plt.ylabel("Tipo beneficio")

plt.title("Cantidad de mujeres por beneficio")

plt.show()
```

Para hombres:

```
f_beneficio= sns.countplot(data=data[data['Sexo']=='M']['Beneficio'])

plt.xlabel("Cantidad de Hombres ")

plt.ylabel("Tipo beneficio")

plt.title("Cantidad de Hombres por beneficio")

plt.show()
```

Paso 4

Luego procedemos a introducir un código que nos indique, en qué universidades se encuentran tanto las mujeres como los hombres

Para mujeres:

```
f_universidad= sns.countplot(data=data[data['Sexo']=='F']['Entidad Organización'])  
  
plt.xlabel("Cantidad de mujeres matriculadas")  
  
plt.ylabel("Universidades")  
  
plt.title("Cantidad de mujeres matriculadas por universidad")
```

Para hombres:

```
f_universidad= sns.countplot(data=data[data['Sexo']=='M']['Entidad Organización'])  
  
plt.xlabel("Cantidad de hombres matriculados")  
  
plt.ylabel("Universidades")  
  
plt.title("Cantidad de hombres matriculadas por universidad")
```

Paso 5

Luego queremos saber a qué nivel educativo ingresan estas personas. Si ingresan a técnicas, tecnologías, pregrados etc. Entonces introducimos el código correspondiente para que se visualice en un diagrama de torta.

Para mujeres:

```
clasificacion_academica_f = data[data['Sexo'] == 'F']['Clasificacion  
Academica'].value_counts()
```

```
plt.figure(figsize=(8, 8))

plt.pie(clasificacion_academica_f, labels=clasificacion_academica_f.index,
autopct='%1.1f%%', startangle=90)

plt.title("Distribución de clasificación academica para mujeres")

plt.show()
```

Para hombres:

```
clasificacion_academica_m = data[data['Sexo'] == 'M']['Clasificacion
Academica'].value_counts()

plt.figure(figsize=(8, 8))

plt.pie(clasificacion_academica_m, labels=clasificacion_academica_m.index,
autopct='%1.1f%%', startangle=90)

plt.title("Distribución de clasificación academica para hombres")

plt.show()
```

Paso 6

Posteriormente surge entonces la duda: ¿de qué lugares vienen la mayoría de estudiantes?.

Para tal propósito introducimos el siguiente código

```
sns.countplot(x='Municipio', data=data)

plt.xlabel("Cantidad estudiantes por municipio ")

plt.ylabel("Municipio")
```

```
plt.title("Cantidad de estudiantes por municipio")

plt.xticks(rotation=90)

plt.show()
```

Esta información también la desagregados por géneros.

Paso 7

Finalmente deseamos hacer un gráfico 3D que relacionara 3 variables: Incentivo, sexo y ciudad. Para tal propósito introducimos los siguientes códigos:

```
# prompt: hazme un grafico 3d para las variables sexo, municipio y beneficio

from mpl_toolkits.mplot3d import Axes3D

data['Sexo_code'] = data['Sexo'].astype('category').cat.codes

data['Municipio_code'] = data['Municipio'].astype('category').cat.codes

data['Beneficio_code'] = data['Beneficio'].astype('category').cat.codes

fig = plt.figure(figsize=(12, 10))

ax = fig.add_subplot(111, projection='3d')

ax.scatter(data['Sexo_code'] + np.random.rand(len(data)) * 0.2 - 0.1,

           data['Municipio_code'] + np.random.rand(len(data)) * 0.2 - 0.1,

           data['Beneficio_code'] + np.random.rand(len(data)) * 0.2 - 0.1,
```

```

c=data['Sexo_code'], cmap='viridis', marker='o') # Usar el

ax.set_xlabel('Sexo')

ax.set_ylabel('Municipio')

ax.set_zlabel('Beneficio')

ax.set_xticks(data['Sexo_code'].unique())

ax.set_xticklabels(data['Sexo'].unique(), rotation=45, ha='right')


ax.set_yticks(data['Municipio_code'].unique())

ax.set_yticklabels(data['Municipio'].unique(), rotation=-45, ha='left')


ax.set_zticks(data['Beneficio_code'].unique())

ax.set_zticklabels(data['Beneficio'].unique())


plt.title('Gráfico 3D de Sexo, Municipio y Beneficio')

plt.tight_layout()

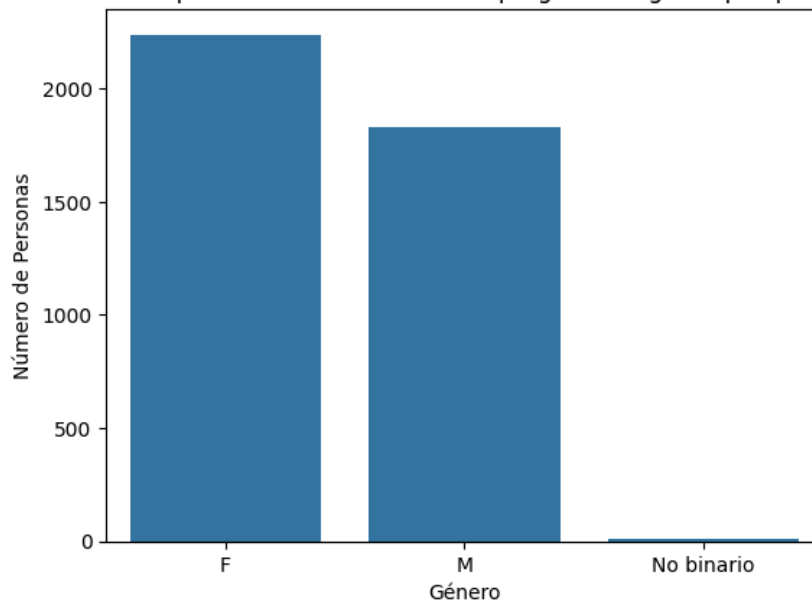
plt.show()

```

5. Hallazgos

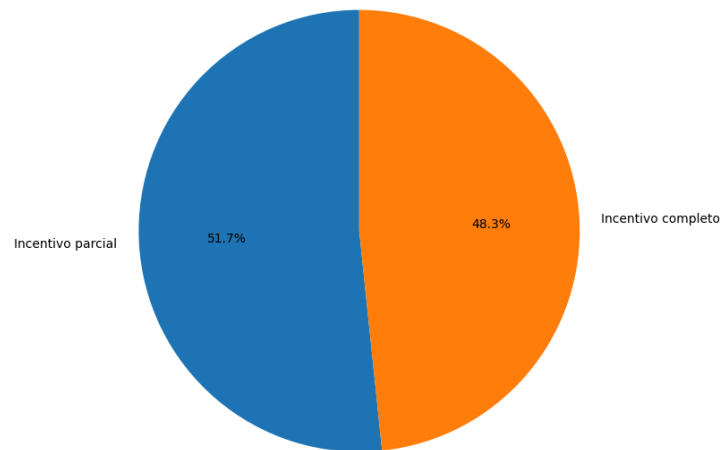
Se observa en esta primera gráfica que el programa ha beneficiado significativamente a las mujeres con un total de 2338 mujeres beneficiarias. Por su parte 1828 hombres fueron beneficiados mientras que 13 personas de género no binario recibieron el apoyo.

Cantidad de personas beneficiarias del programa DigiCampus por genero

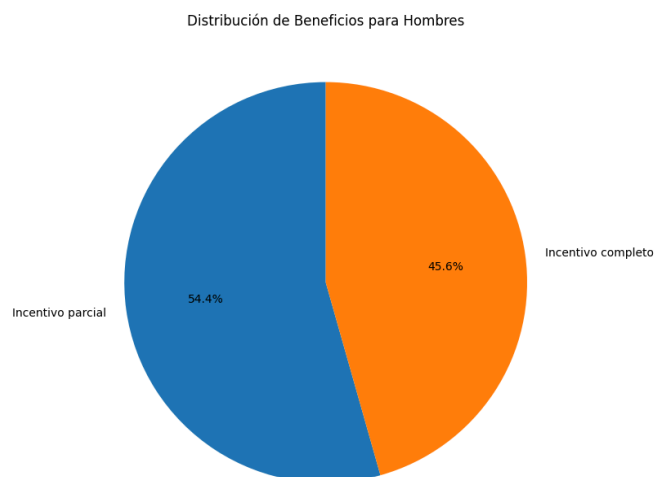


Distribución de los beneficios

Distribución de Beneficios para Mujeres

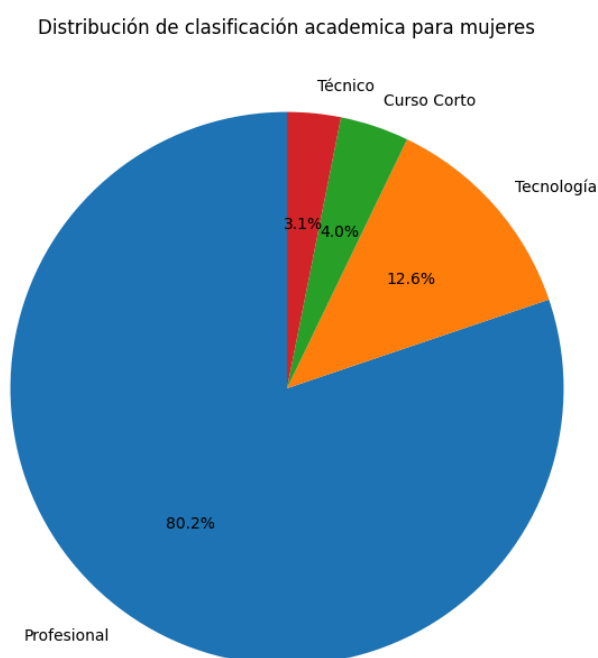


Se observa en el gráfico que el 51% del total de las mujeres beneficiadas recibieron un incentivo parcial, mientras que un 48% de ellas recibió incentivo completo.



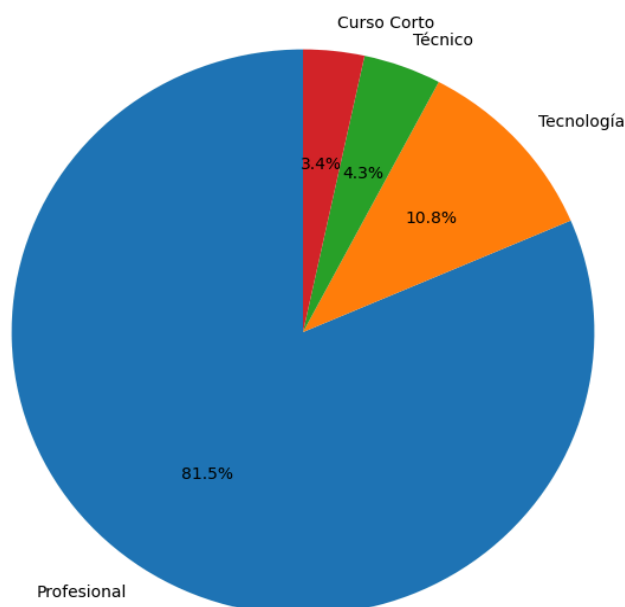
Por parte de los hombres los resultados no varían mucho con respecto a las mujeres. 54,4% de los hombres recibieron incentivo parcial mientras que el 45,6% recibieron incentivo completo

Clasificación académica



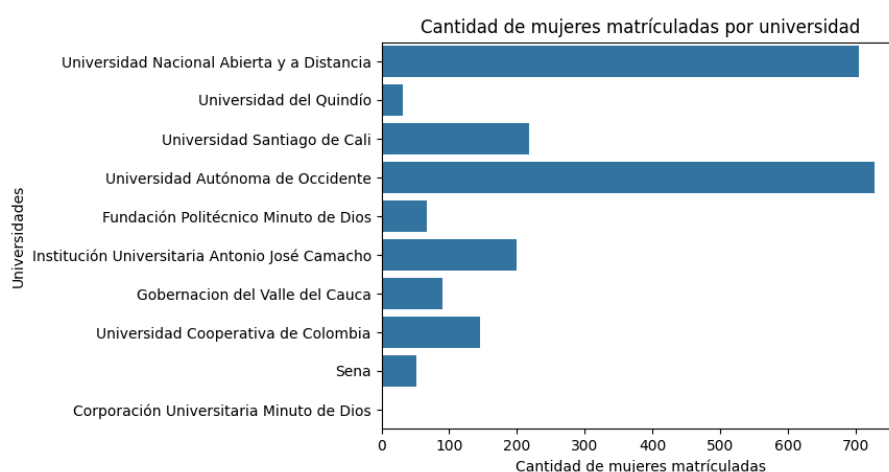
Con respecto a los niveles académicos a los que las mujeres han ingresado, se puede observar que su mayoría (80,2%) ingresaron a una educación de nivel profesional, mientras que 12.6% ingresaron a tecnología, 4% a curso corto, y por último 3,1% a nivel técnico.

Distribución de clasificación académica para hombres



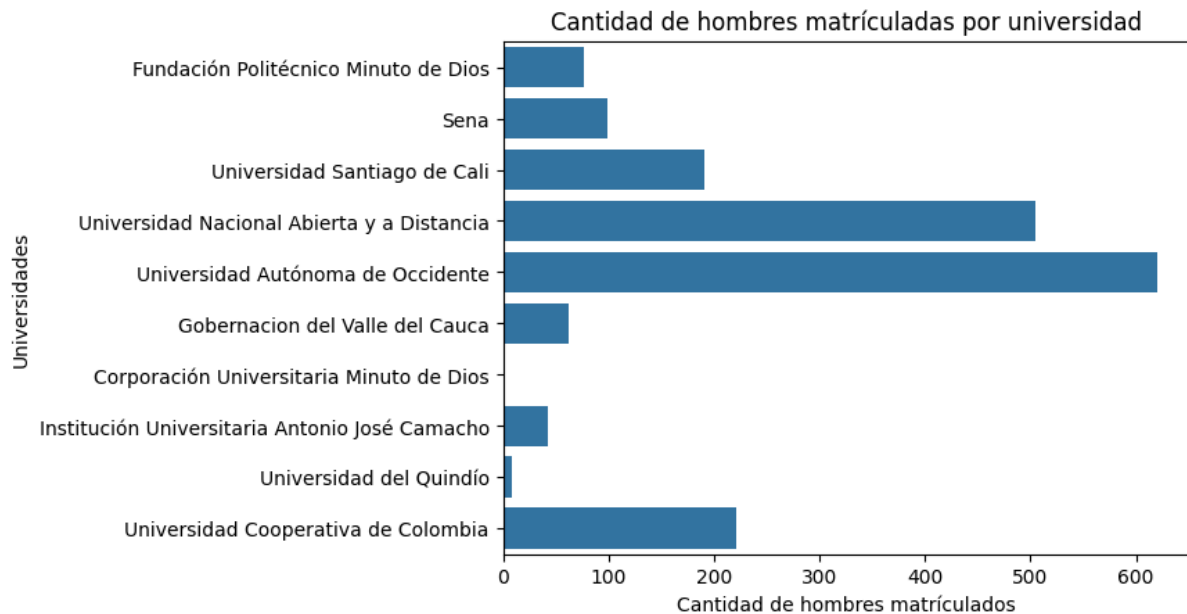
Reflejando un poco la misma dinámica de las mujeres, la mayoría de los hombres (81,5%) ingresaron a una carrera profesional, mientras que 10,8% a tecnología, 4,3% a un curso técnico y 3,4 a cursos cortos.

Instituciones a las que ingresan



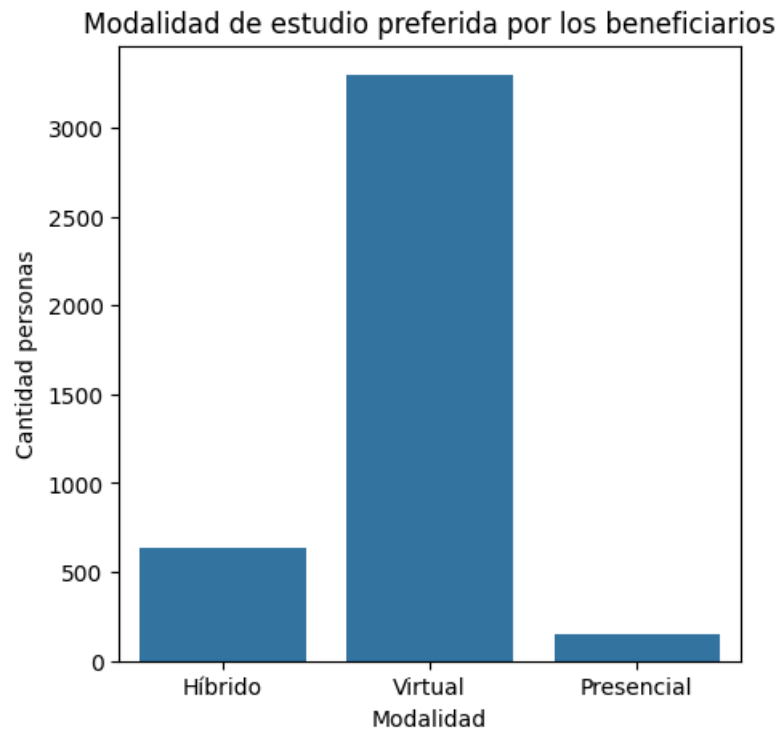
Con respecto a las instituciones a las que las mujeres han ingresado con el beneficio se observa que la Universidad Nacional Abierta y a Distancia ha sido la más preferida por las mujeres en compañía con la Universidad Autónoma de Occidente. Ambas con una cantidad

de aproximadamente 700 mujeres. Por su parte la Universidad del Quindío ha sido la menos apetecida por este sector poblacional.



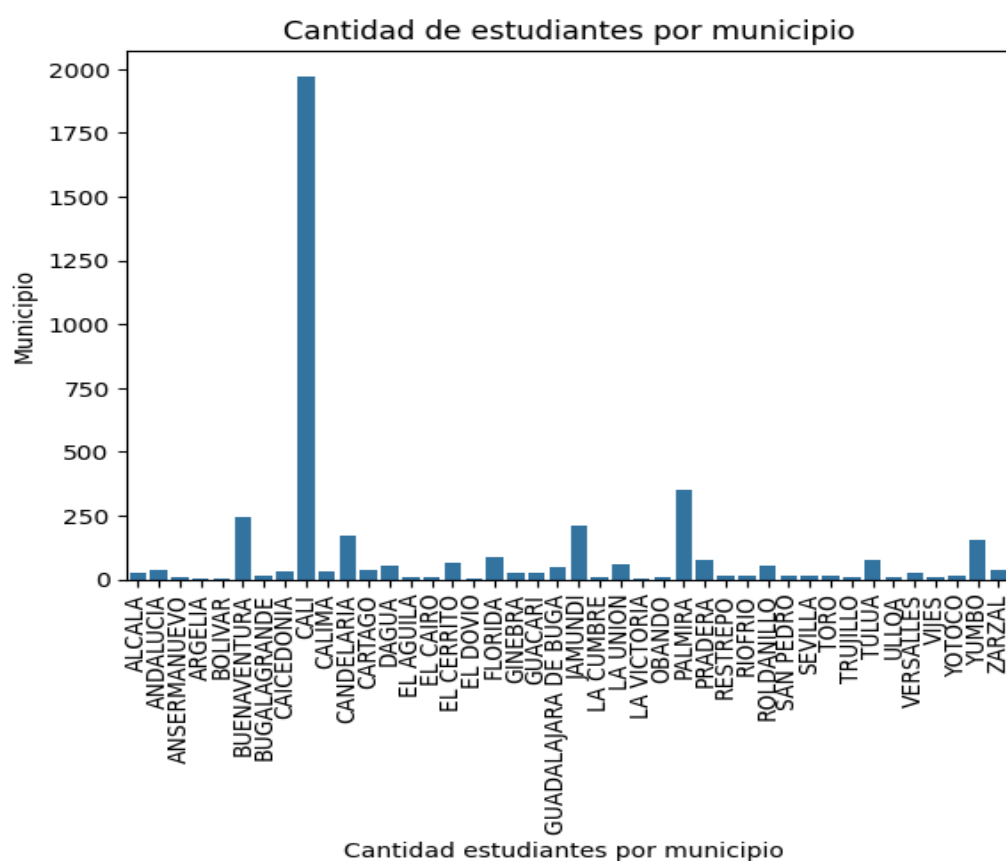
Los hombres, reflejando un comportamiento similar al de las mujeres, han preferido primeramente la Universidad Autónoma de Occidente y seguidamente la UNAD. En comparación con las mujeres, la Universidad Cooperativa de Colombia ha sido más apetecida por los hombres en comparación con la Universidad de Santiago de Cali.

Modalidad de estudio



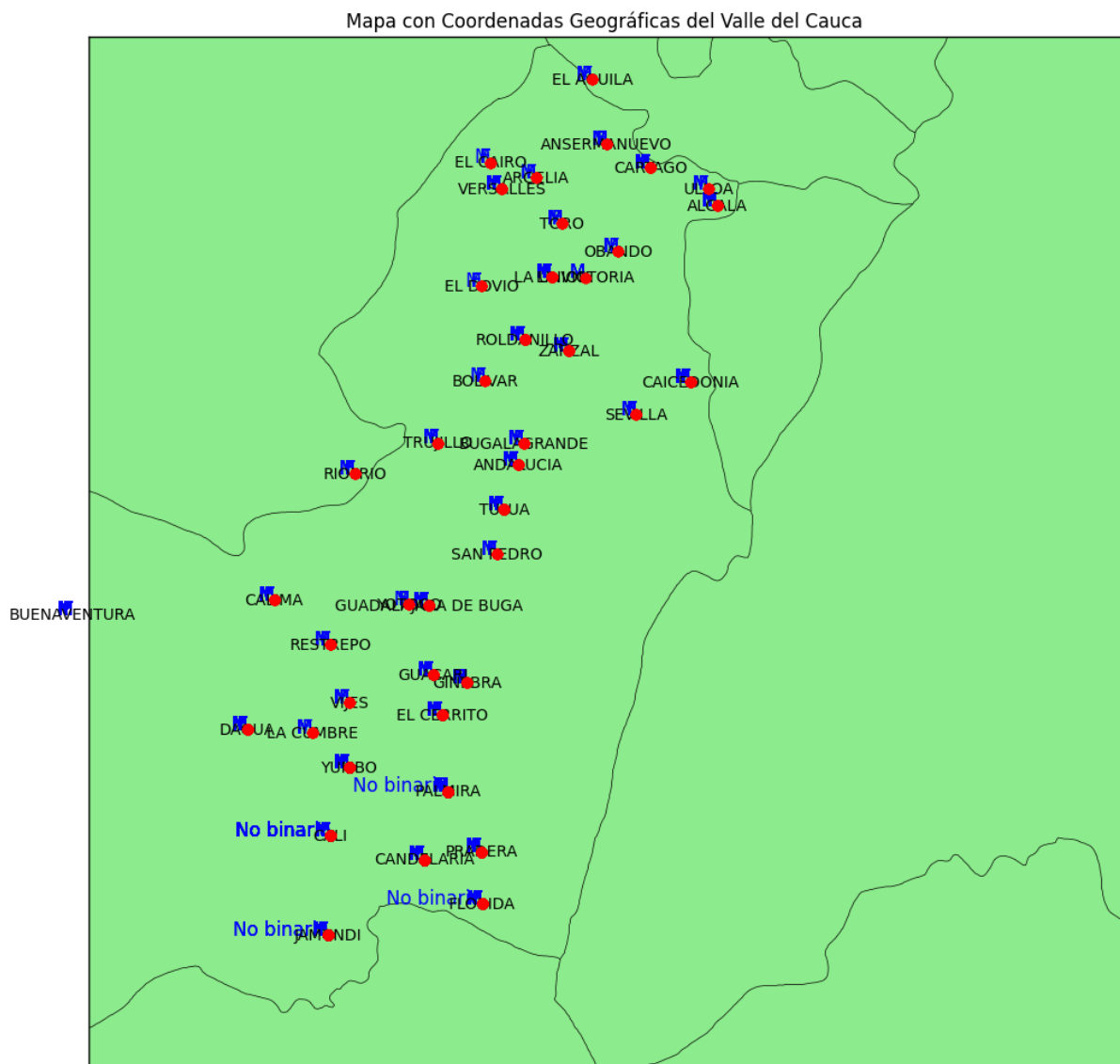
En total la mayoría de la población beneficiaria ha preferido notoriamente la modalidad de estudio virtual. Lo que indica un impacto positivo de la política pública en términos de inserción estudiantil a distancia.

Municipios de donde provienen la mayoría de estudiantes



Por su parte, se observa en la gráfica que la mayoría de los beneficiarios provienen de la ciudad de Calí, seguido por Palmira y Buenaventura.

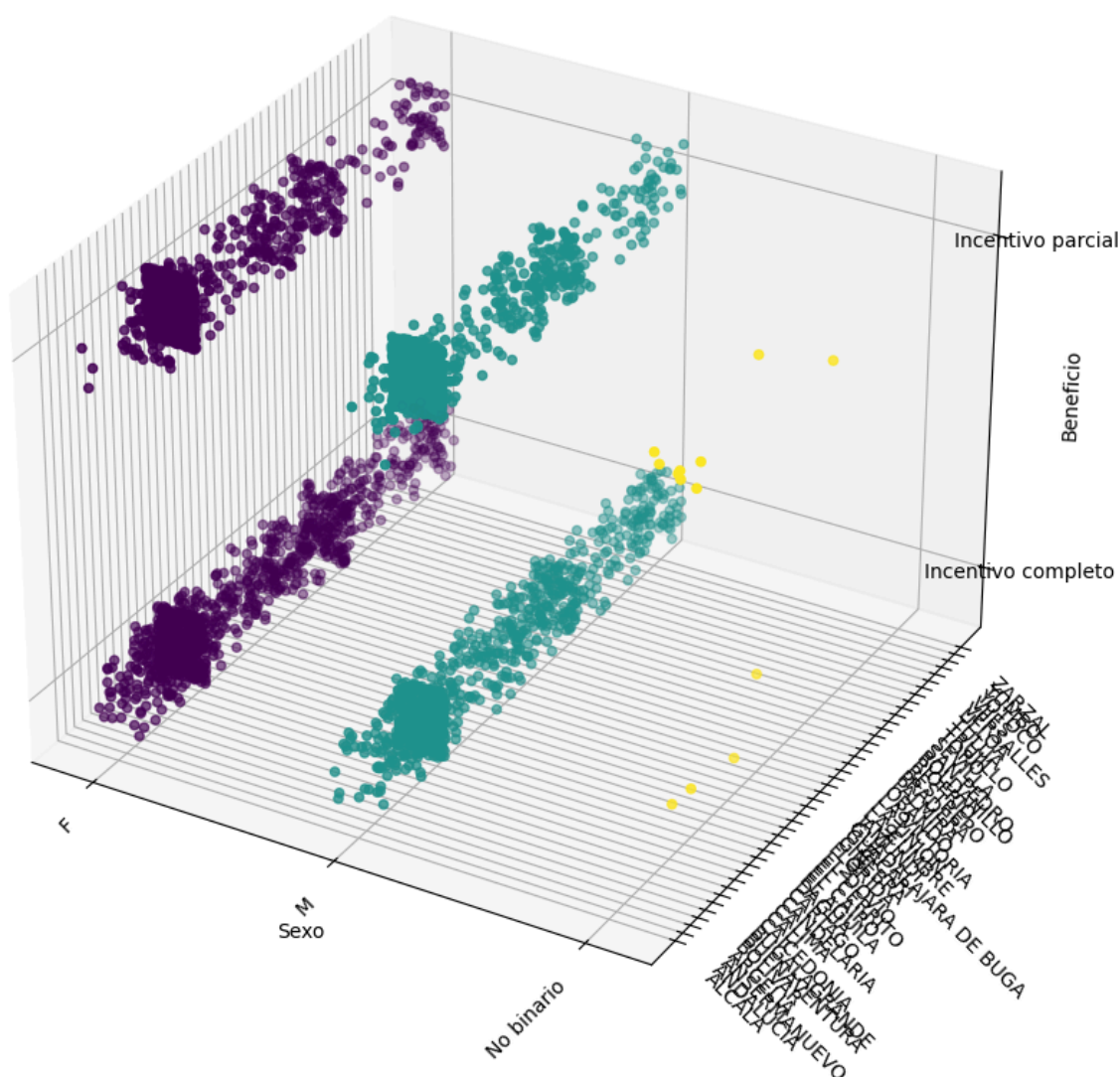
Alcance de la política



A pesar de que, como se observó con antelación, la mayor parte de los beneficiarios provienen de Calí, se observa que los beneficiarios del programa están dispersos en el Departamento del Valle. Lo que da indicios de que la política pública busca ser inclusiva

Relación incentivo, Género y Ciudad.

Gráfico 3D de Sexo, Municipio y Beneficio



En hilo con los gráficos anteriores, se observa que la mayor densidad de la población que recibe el beneficio se encuentra en el género Femenino y Masculino en incentivos completos y parciales distribuidos similarmente. Se nota además que la mayoría de participantes provienen de 3 ciudades.

Conclusiones

1. El programa DigiCampus ha logrado un alcance significativo, beneficiando a más de 4000 personas y mostrando señales de inclusión tanto de género como territorial.

2. La distribución de becas entre hombres y mujeres es relativamente equitativa, sin indicios fuertes de sesgo de género.
3. La virtualidad ha sido clave para la implementación exitosa del programa, lo cual sugiere que cualquier mejora futura debería seguir fortaleciendo la infraestructura digital en zonas rurales y vulnerables.
4. La metodología empleada fue adecuada para el análisis propuesto: se identificaron patrones útiles mediante limpieza de datos, visualización y análisis estadístico básico.
5. El proyecto sienta bases importantes para futuras investigaciones, no solo sobre DigiCampus, sino sobre otras políticas públicas similares que busquen mejorar el acceso equitativo a la educación superior.
6. Se puede indicar que el impacto de la política pública ha alcanzado hasta ahora sus objetivos del fomento de la educación en la población joven.

Reflexiones

- a) Importancia de los datos para evaluar políticas públicas:
 - Este trabajo muestra cómo el análisis de datos puede ofrecer una visión profunda sobre la efectividad y el alcance de un programa como DigiCampus.
 - El uso de herramientas como Python, pandas y seaborn evidencia cómo el conocimiento técnico puede ser aplicado para resolver preguntas sociales relevantes.
- b) Accesibilidad e inclusión territorial:
 - Aunque la mayoría de beneficiarios provienen de Cali, se observa una distribución amplia en el departamento del Valle del Cauca, lo que refleja un intento por parte del programa de ser territorialmente inclusivo.
 - El predominio de la modalidad virtual sugiere una adaptación a las condiciones de conectividad y disponibilidad de tiempo de los estudiantes.
- c) Diferencias en elección de instituciones y modalidades educativas:
 - Las universidades más escogidas varían ligeramente según el género, pero reflejan preferencias comunes por la educación a distancia.

- La mayoría de beneficiarios optan por programas profesionales, lo que habla de aspiraciones formativas altas dentro de la población beneficiaria.

References

Ministerio de las TIC. (2025, 01 09). *Apropiación TIC del Valle del Cauca*. Datos abiertos.

https://www.datos.gov.co/en/dataset/Apropiaci-n-TIC-del-Valle-del-Cauca/atu5-g8g5/about_data